International Journal of Computer Science Engineering
and Information Technology Research (IJCSEITR)
ISSN(P): 2249-6831; ISSN(E): 2249-7943
Vol. 7, Issue 2, Apr 2017, 1-10

SMART MOTIF DISCOVERY ON ECG SIGNAL USING AUTOCORRELATION AND VARIABLE THRESHOLD

HARJEET KAUR[1] & RAJINDER KAUR[2]

[1]*Research Scholer, Chandigarh University, Gharuan, India*

[2]*Assistant Professor, Chandigarh University, Gharuan, India*

ABSTRACT

The ECG signal is the prominent solution to determine the heart beat rate of the person to evaluate the health of heart. The telemedicine or body area networks have enabled the remote live ECG monitoring of the patients staying at their homes or remote areas. The device used to sense the ECG signal is called Holter, and is a wearable device. The Holter collect the ECG signal from the human body and forwards it towards the healthcare database. Then the data is evaluated on the cloud healthcare server to determine the heart beat rate obtained from the signal by using the motif discovery method. The motif discovery method evaluates the signal and finds the appropriate and valid heart beats to determine the beat rate along with the abnormalities in the signal to examine the patient's critical health report. In case the heart beat rate is too high or contain number of abnormalities, the alarm is raised towards the concerned physician and the patient to take the preventive measure. In this paper, the flexible length motif discovery algorithm has been designed for the variable motif detection. The proposed model is used to detect the motifs in the input ECG signal. The motif discovery is performed between the given minimum and maximum motif length. The experimental results have proven the efficiency of the proposed model in detecting the motifs in the input signal.

KEYWORDS: Motif Detection, Variable Length Pattern Discovery, Precise Pattern Discovery & ECG Signal Analysis

Received: Jan 07, 2017; **Accepted:** Feb 11, 2017; **Published:** Feb 16, 2017; **Paper Id.:** IJCSEITRAPR20171

Original Article

INTRODUCTION

The body area networks and health community in hospital are considered commercial interests in optimization, classification, compression or de-noising of Electrocardiogram signals. The Electrocardiogram is most general known to recognized and used biomedical signal. Most important fact which the ECG waveform can be altered by cardiovascular abnormalities and diseases such as arrhythmia, etc. The heart has four chambers and various one-way valves. Septum is divided the heart into two parts. The opposite side of the heart receives the oxygenated blood from pumps andlungs it into systemic loop for distribution throughout the body.

On the other side, a specialized conduction system and Co-ordinated electrical events of heart play main roles in rhythmic contractile activity of heart. Atrio-Ventricular (AV) nodes and the Sino-Atrial (SA). SA (Sino Atrial) node is a native. The SA (Sino Atrial) pacemaker cells can be generate action potentials at 60-80 times per minute can take over the role if for some reason the SA node fails. Some analyzers assist the doctor by producing ainterpretation; others provide a fixed number of parameters by which the doctor can make his diagnosis.

The data acquisition is aninitial step in ECG signal analysis. The data is collected from real subjects but in this paper it is gathered from online database. Next step is signal pre-processing. In this step retrieve ECG signals can be filtered to remove the noise and to apply optimization or compression. The third step is QRS detection that corresponds to period of ventricular depolarization or contraction of heart. Fourth step is to find smallest set of features which maximize the classification performance of next step. Final step is classification of signal into three different cardiac conditions. Many different techniques like as artificial neural network, statistical classifiers and artificial intelligence can be used for ECG classification. Basically, the ECG is one of oldest and most well-known instrument-bound measurements in medical applications. It hassucceed the progress of instrumentation technology.

LITERATURE REVIEW

Harvey, Dustin Y. et. al. [1] has deployed on automated feature design for numeric sequence classification by genetic programming. In this paper, intended a Auto fead genetic programming method for automatic feature design. In this method, a genetic programming alternative evolves a population of candidate features built from a library of sequence-handling functions. Fedotov, Aleksandr A. et. al. [2] has worked on the uncertainties in measurement of nonlinear dynamics in heart rate variability. Harvey, Dustin Y. et al. [3] has performed robust evaluation of time series classification algorithms is used for structural health monitoring. Info-gap decision theory indicates the non-probabilistic evaluation of the robustness of competing models and these systems that are given variety of decision making applications.

Wang, Jin et al. [4] has projected towards bag-of-words represented for biomedical time series classification. The biomedical time series is then described as a histogram of code words, each and every input of which is the count of a code word sconduct in the time series classification. Although the temporal sequences of the local segments are forgotten, the bag-of-words represented as able to capture high-level structural data since both local and global structural information data are well handled. Xing, Zhengzhenget. al. [5] has performed a brief survey on sequence classification. This makes sequence classification a more demanding or interesting task than classification on feature vectors. Yahyaoui, Hamdiet. al. [6] has developed the feature-based trust sequence classification algorithm. By analyzing the available related natures of services, trust patterns are defined to explain trust sequences based on three criteria: its overall nature, the starting nature and ending nature.

EXPERIMENTAL DESIGN

In this paper, ECG signal has been optimized with use of particle swarm optimization (PSO) and genetic algorithm (GA) for quick response motif detection. Genetic algorithm is employed because of effectiveness. Our study has proved that particle swarm optimization and genetic algorithm are the good signal optimization algorithms. The genetic algorithm gave a multiple choice is made because it is more impressive and consistent than particle swarm optimization. We have also implemented the particle swarm optimization (PSO) algorithm for the FECG signal for optimization, which is used for the comparison of both of the algorithms. The optimization process decrease the load on the data communication links; hence increase the detection rate of the ECG signal motifs. The optimization performance and convergence speed of regeneration quality maintenance is evaluated by calculating the performance specification like mean squared error (MSE) and peak signal to noise ratio (PSNR) etc.

The signal quality has been measured using various performance parameters being mean square error and peak signal to noise ratio.

$$PSNR = 10.\log_{10}\left(\frac{255*255}{MSE}\right) \quad (1)$$

$$\mathrm{MSE} = \frac{1}{\mathrm{n}\sum_{\mathrm{i}=0}^{\mathrm{n}}(\mathrm{X}-\mathrm{X}')^2} \quad (2)$$

Where X is the Original ECG data and X' is the transformed ECG data and n is the number of samples. The detail of our implementation of GA is described as follows:

Algorithm 1: Genetic Algorithm

Input: Original ECG Signal

Output: Optimized ECG Signal

Step 0*: parameter initialization (pop size, mutation rate and max pop generations)*

Step 1*: Initial population is created anyway*

Step 2*: chromosome fitness evaluation/calculation*

Step 3*: If ending condition is not satisfied do [step 4-8]*

Step 4*: execute selection from P(t) to P(t+1)*

Step 5*: perform mutation and crossover*

Step 6*: recomputed the chromosomes fitness*

Step 7*: t=t+1*

Step 8*: go to step 3*

The described as following is the implementation PSO in our implementation

Algorithm 2: Particle Swarm Optimization

Input*: Original ECG Signal*

Output*: Optimized ECG Signal*

Step 1*: For every particle*

Step 2*: Initialize the particle formation*

Step 3*: End*

Step 4*: Do (overall)*

Step 5*: Do For every particle perform step 6-8*

Step 6*: Calculate the fitness of all particles*

Step 7*: If fitness value better than personal best (pBest) in the current computational history*

Step 8*: set the pBest equals thegBest*

Step 9*: End*

Step 10*: Choose the best pbest as the gBest for overall solution*

Step 11*: Calculate velocity of the particles according to the equation (i)*

Step 12*: Shift the* particle's *positions according to equation (ii)*

Step 13*: End*

After each iteration Particle swarm optimizer updates velocity and positions of each value in space with following equation (i) and (ii).

$$\mathrm{Q[i]} = \mathrm{q[i]} + \mathrm{CL1} * \mathrm{rand(\,)} * (\mathrm{pB[i]} - \mathrm{pR[i]}) + \mathrm{CL2} * \mathrm{rand(\quad)} * (\mathrm{gB[i]} - \mathrm{pB[i]}) \qquad \textbf{(i)}$$

$$\mathrm{pR[i]} = \mathrm{pR[i]} + \mathrm{q[i]} \qquad \textbf{(ii)}$$

Where q[i] is refers the particle velocity, pR[i] determine the current solution for particle holds randomly. pB[i] and gB[i] are local and global best fitness values for every particle respectively . rand () is a random number between (0,1). CL1, CL2 are the two learning factors.

The brute force algorithm has compared the all possible combination of subsequences of all possible lengths.

Afterwards, the proposed model algorithm computes the product of T1;m with Ti; m is computed and saved in xyi. The proposed model shows the main three loops of the algorithm. Reasonably, the loop in line 17 determine the closest nearest of Ti; j . Lines 12-15 count the dot products for Ti;j given we have the dot products for Ti□1;j in prev XY and store that behind in prev XY for future use (line 15). Note that, all the arrays are of the equivalent size as the time series. The proposed model also includes the distance values of the combination in a list to return. Finally, it count the nearest neighbor of Ti; jis found, we can update or try the best combination for the length of j. And finally when all of the loops are executed, we can output the bests for all lengths.

Algorithm 3: Smart Motif Discovery using the TLV Method

Input:*{max motif length, time series signal}*

Output:*{Motif length, Motif position}*

- Acquire the input ECG signal
- Get the size of the ECG signal in the number of rows and columns
- Initialize the value x_0 to 0
- Initialize the value of xx_0 to 0
- Initialize the window size to minimum motif length
- **Run the iteration equals the total length variance (TLV)**
- **Run the iteration equals the signal length minus window size**
    - Acquire the signal according to the signal lengths
    - Extract the signal values fetched according to the window size

- Apply the sliding window analysis to find the fittest pattern in the input window
- Find the signal pattern in the given data not exceeding the TLV
- If the pattern is found
  - Mark the position of the detected motif
  - Obtain the motif length
  - Add the position to the position array
  - Add the motif length to the length array
  - compute the squared distance to classify the pattern
- Update the best pair for length j if necessary
- Increment the window size by one
- If it is not the last iteration
  - GOTO step 7(a)
- Otherwise
  - Break the iteration
- Output the best pair for all the lengths
- *Return the motif length and position List*

Conventional ECG systems require the patient's presence at the neighboring hospital for the ECG checkup, normal ECG constitute a 10-Electrode placed at a variety of sensitive areas of patient's body for examining the electrical activity of heart. The output of system is depicted onto a graph paper, which gives a crystal clear look for the doctor. Considering the data being presented by ECG machine direct that are executed no defect data. Conventional System gives rapid time input, this cannot be utilize for the analyzing the activity of heart above a large interval. For the ECG data obtained for the Holter the initial step is the noise removal process which undergoes the proposed signal processing algorithm. There will be three essential types of noises will be eradicate which can be most likely caused during the signal recording, signal propagation, etc. Initially we need to abolish the baseline drift which is a type of noise got produce during the signal recording process. A moving average filter will help in the elimination of baseline drift. Removal of the signal burst by use of Savitzky-Golay Filtering is the subsequent step. Following the removal of the signal burst, the subsequent purpose is to eliminate the salt and pepper noise or spikes from the ECG data that arises because of any voltage fluctuation using Median Filter.

Algorithm 4: Overall Signal Processing for the Motif Discovery

- Acquire the ECG signal and read the ECG data in the waveform
- Convert the input data into the vectorized form
- Compute the signal drift using the drift error evaluation

    - Returns the drift vector equals the length of ECG signal length
    - Remove the drift from the original signal
    - Return the signal after the drift removal
- Evaluate the degree of fitness over the signal after drift removal
- Run the iteration equals the number of blocks in the input signal
    - Input the degree of polynomial up to the fourth order
    - Compute the least squares method
    - Compute the polynomial value
    - Obtain the fittest polynomial coefficients
    - Shift each signal block according to the computed value
    - If it is the last iteration
        - Return the signal
- Analyze the R-peaks using the early stage detection using the relative gradient in the block based manner
- Analyze the signal coefficients {P, Q, R, S, T} of each curve of the motif
- Determine the RR interval between the two R-peaks
- Identify the first R-wave in the signal
- Enlist and mark each R-wave after the classification of the first R-wave
- Finish the iteration after identifying the last R-wave using the decision rule
- Identify the false negative cases
- Delineate the sub-signal region in the false negative case by employing the change of slope rate for false positive case elimination
- Apply the median filter once again
- Remove base line drift from the final signal.

The exact or probable motif discovery methods are utilized for the pattern discovery practices. The variety of noise data interrupts the accuracy of the motif discovery methods. The noise like electrode motion, baseline wandering, power-line interference based noise and muscle noises. There are several methods already proposed for the enhancement in accuracy of the motif discovery by normalizing the signal noise applications with adaptive solutions in order to overcome the complexities of the motif discovery methods. Also the signal length produces the major problem as it increases the execution time for the motif discovery. The popular motif discover methods utilize the linear and non-linear filtering and transformations along with the decision rule discovery based motif finding algorithm. In order to find the stronger motifs the stronger edges are discovered as the edge after applying the differential operations. The hibert

transform is also utilized for the purpose motif normalization for the purpose of motif discovery.

Algorithm5: Motif Discovery of One Training and Test Motifusing Autocorrelation and Variable Threshold

- Input the signal data with the length of N-entities denoted X(t)
- Input the step size between the range of 0.04 seconds to 0.12 seconds.

  *Motif_step_size=step_size*fs*

  Where the Motif_step_size decides the time interval for the motif discovery, step_size defines the time interval range for the recording of the signal, fs denotes the frequency of the signal values per seconds.

- Input the initial threshold value (*th*) for the motif discovery.
- Find the initial motif according to the input threshold (*th*)
- Find the numeric sequence within the given range by defining the alternative range formed from threshold
- Finalize the training motif length according to the initial pattern discovery
- Discovery the maximum amplitude pattern in particular window and assign the pattern to the training motif data.
  - For*i*=1:maxLength
    - If (*th*(*i*)>max(*i*))
    - If ((*th*(*i*)>max(*i*)) && (*th*(*i*)<max(*i*-1)))
    - *Thr(i)*=motifData(cnt)
    - Increment cnt
  - Otherwise
    - If ((*th*(*i*)>max(*i*)) && (*th*(*i*)>max(*i*-1)))
    - *Thr(i)*=motifData(cnt)
    - Increment cnt
  - Repeat i(a)
- Find autocorrelation between the signal and training motifs For*j*=1:length(motif_step_size)

  For*i*=1:(length(*X*)-*motif_step_size*(*j*))*motif(j)*=*sum*(*j*)+abs(*X(i)*)*abs(*X*(*i*+*motif_step_size*(*j*)));

  motif(j)=*sum(j)*/((length(*X*))-(*motif_step_size*(*j*)));

- Return the discovered motif data matrix and motif location map.

RESULTS ANALYSIS

The result analysis has been performed at the ECG data collected from the Physionet MIT database. The ECG signal for the length of 90 seconds has been gathered in each file. The data of six randomly selected patients has been evaluated for the motif discovery of the ECG data. The proposed model has been designed to work with the signal length

of maximum ten thousands. The elapsed time for the motif discovery has been recorded in the milliseconds for each entry in the following table 1:

Table 1: Elapsed Time by Smart Motif Discovery Algorithm Over the Physionet ECG Database

Data Files	Total Time for Motif Discovery (Signal Length=10000)
Motif 1	23. 86
Motif 2	15. 63
Motif 3	40. 22
Motif 4	26. 79
Motif 5	34. 29
Motif 6	18.44
Motif 7	21.48

Table 2: Recorded Performance Parameters (Statistical Errors)

Patient ID	Motif Classification Before Optimization			Motif Classification After Optimization		
	PSNR	MSE	PRD	PSNR	MSE	PRD
Motif 1	51.9469	0.4153	98.3802	64.4333	0.0234	82.6978
Motif 2	50.0849	0.6377	97.0949	60.3637	0.0598	83.5218
Motif 3	50.0708	0.6397	97.1725	60.4399	0.0588	83.5399
Motif 4	50.1901	0.6224	97.0292	60.4795	0.0582	83.6378
Motif 5	50.1850	0.6231	97.0093	60.4421	0.0587	83.6450
Motif 6	50.3392	0.6014	98.4157	60.7085	0.0552	84.3760
Motif 7	50.3899	0.5944	98.1588	60.8683	0.0532	83.9272

The results in table 2 have proved that particle swarm optimization performed better than Genetic Algorithm. PSO has taken almost less than half energy than GA for the ECG signal optimization. The energy of ZigBee environment has also been recorded to propagate the ECG signal of each patient. Each patient's ECG signal examined under this project is of the length of 90 seconds and has recorded on the 356 samples per second.

The results of the elapsed time have been collected from the experiments performed over the given ECG data collected for the various patients. The algorithm has handled all of the essential processes which include the training motif and testing motif discovery are computed for the time domain in this implementation. The following table describes the discovered motif lengths and the time for the specific motif discovery over the EEG database obtained form the Moen's motif discovery portfolio. The variable length motifs has been discovered from the input signal for each motif, the result analysis has been recorded and published in the following table 3:

Table 3: Variable Length Motif Discovery in the EEG Signal

Index of Motifs	Motif Length	Elapsed Time (In Seconds)
1	127	0.021123
2	142	0.020373
3	157	0.019857
4	163	0.019831
5	138	0.01958
6	172	0.020239
7	253	0.020242
8	221	0.020299
9	162	0.021009
10	181	0.020092

The above table 3 describes the results obtained from the motifs discovery algorithm while executed to discover the motifs between the length of 64 and 256. The above results show that the proposed algorithm has taken almost similar time to detect the motifs in the given signal.

CONCLUSIONS

At First, in this paper, the major focus has been on discovering the normal and abnormal motifs using the floating threshold with minimized motif deterministic templates. All of the experiments have been conducted over the ECG data recorded with sampling rate of 360 samples per second to record digital ECG signal with 11-bit resolution using the 10 Millivolts range. The semi-supervised proposed (L,D)-motif discovery method is being developed as the quick response method with high precision to handle the maximum number of patients in the pool over the high-density cloud healthcare monitoring service. The (L,D)-motif discovery method works on the collaborative ECG signal obtained from various patients together, where L is the length of each signal and D denotes the dimensions (equals number of active patients). Second, the ECG signal handling is discussed in this paper using the particle swarm optimization (PSO) and genetic algorithm (GA) to optimize the signal. The performance evaluation of the proposed model solutions using GA and PSO have been obtained as elapsed time, precision and signal quality measure (PSNR).

REFERENCES

1. *Harvey, D. Y., & Todd, M. D. (2015). Automated feature design for numeric sequence classification by genetic programming. IEEE Transactions on Evolutionary Computation, 19(4), 474-489.*
2. *Fedotov, A. A., Akulova, A. S., & Akulov, S. A. (2015). Uncertainties in measurement of nonlinear dynamics in heart rate variability. In 6th European Conference of the International Federation for Medical and Biological Engineering (pp. 102-105). Springer International Publishing.*
3. *Harvey, D. Y., Worden, K., & Todd, M. D. (2014, March). Robust evaluation of time series classification algorithms for structural health monitoring.InSPIE Smart Structures and Materials+ Nondestructive Evaluation and Health Monitoring (pp. 90640K-90640K).International Society for Optics and Photonics.*
4. *Wang, J., Liu, P., She, M. F., Nahavandi, S., & Kouzani, A. (2013). Bag-of-words representation for biomedical time series classification. Biomedical Signal Processing and Control, 8(6), 634-644.*
5. *Xing, Z., Pei, J., & Keogh, E. (2010). A brief survey on sequence classification. ACM SIGKDD Explorations Newsletter, 12(1), 40-48.*
6. *Yahyaoui, H., & Al-Mutairi, A. (2016). A feature-based trust sequence classification algorithm. Information Sciences, 328, 455-484.*
7. *Chen, L., & Guo, G. (2015). Nearest neighbor classification of categorical data by attributes weighting. Expert Systems with Applications, 42(6), 3142-3149.*
8. *Gabr, M. M., & Fatehy, L. M. (2013). Time series classification. Journal of Statistics Applications & Probability, 2(2), 123-133.*
9. *Chaturvedi, B., & Patil, N. (2015, June). A novel semi-supervised approach for protein sequence classification.In Advance Computing Conference (IACC), 2015 IEEE International (pp. 1158-1162).IEEE.*

10. *Shi, Y., &Eberhart, R. (1998, May). A modified particle swarm optimizer. InEvolutionary Computation Proceedings, 1998. IEEE World Congress on Computational Intelligence., The 1998 IEEE International Conference on (pp. 69-73). IEEE.*
11. *Eberhart, R. C., & Shi, Y. (1998, March). Comparison between genetic algorithms and particle swarm optimization. In International Conference on Evolutionary Programming (pp. 611-616).Springer Berlin Heidelberg.*